

BU-150-m

THE USE OF ALTERNATIVE TRAITS IN
SELECTION PROGRAMS

Technical Report No. 11

Department of Navy
Office of Naval Research

Contract No. Nonr-409(39)
Project No. (NR 042-212)

S. R. Searle
Biometrics Unit
New York State College of Agriculture
Cornell University
Ithaca, New York

This work was supported in part by the Office of Naval Research.
Reproduction in whole or in part is permitted for any purpose of the
United States Government.

THE USE OF ALTERNATIVE TRAITS IN SELECTION PROGRAMS^{1,2}

BU-150-M

S. R. Searle
Cornell University

August, 1963

Introduction

Attempts at changing the average genetic merit of a trait in a population by means of a selection program are usually made by selecting on the phenotype of the trait concerned. But the possibility might also be considered of utilizing another trait, through selecting on it rather than on the main trait itself. Whereas simultaneous improvement of two or more traits usually involves a selection program based on the phenotypes of the traits involved, for which several procedures are available (see for example, Young and Weiler, 1960), this paper considers the problem of improving a single trait (hereafter referred to as the basic trait) through selection on other, alternative traits. In many practical situations the use of an alternative trait may, for economic or other reasons, be an attractive prospect as the basis of a selection program -- for example the use of part lactation records in preference to complete lactation records in dairy cattle selection; but regardless of the potency of such reasons, one must first answer the question "how efficient is selecting on an alternative trait compared to selecting on the basic trait?"

One might well want both the heritability and repeatability of the alternative trait to exceed that of the basic trait, and one would presumably desire both the phenotypic and genetic correlations between the two traits to be as close to unity as possible, for a high genetic correlation indicates that on the average an animal selected as having high genetic merit for the alternative trait will also be high in genetic merit for the basic trait; and a phenotypic correlation close to unity indicates that animals selected on their alternative trait phenotypes are likely to also be superior in their basic trait phenotypes. Unfortunately these criteria will not necessarily all be satisfied for an alternative trait that is considered in any particular case. Furthermore, it appears

¹ Research supported in part by the Office of Naval Research, Project No. NR 042-212, Contract No. Nonr-401(39). Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

² To be presented at the Fifth International Biometric Conference, Cambridge, England, September, 1963.

that the only comprehensive criterion available is that suggested by Lerner and Cruden (1948), of assessing the efficiency of an alternative trait by comparing the rates of genetic progress in improving the genetic merit for the basic trait under the two selection programs, through expressing the rate when selecting on the alternative trait as a fraction of that when selecting on the basic trait.

Relative selection efficiency

We will consider the case in which the phenotype of the basic trait is denoted by Y and the corresponding additive genetic merit by y . Heritability in the narrow sense is accordingly $h_y = \sigma_y^2 / \sigma_Y^2$, the ratio of the additive genetic variance, σ_y^2 , to the phenotypic variance, σ_Y^2 . If the selection differential based on selecting according to the phenotypes Y is i_Y , measured in standard deviation units, the expected mean superiority of y in the selected individuals over the mean of the population from which they were selected is $R_{Yy} i_Y \sigma_y$, where R_{Yy} denotes the correlation between Y and y . This is often called the response to selection; it represents the expected rate of genetic improvement (i.e. improvement in y) resulting from selection on the phenotype Y . Likewise, if selection is made on the phenotype of an alternative trait X , for which the selection differential is i_X and the correlation with y is R_{Xy} , then the rate of improvement in y is $R_{Xy} i_X \sigma_y$. Selection for y using X is sometimes referred to as indirect selection (Falconer, 1960, for example) and $R_{Xy} i_X \sigma_y$ as the correlated response. We shall refer to the ratio of the responses as the relative selection efficiency of X , $RSE(X)$, namely the expected rate of improvement in y when selecting on X relative to that when selecting on Y . Hence

$$RSE(X) = \frac{R_{Xy} i_X \sigma_y}{R_{Yy} i_Y \sigma_y} = \frac{R_{Xy}}{R_{Yy}} \left(\frac{i_X}{i_Y} \right) .$$

On assuming that the intensity of selection is the same using X as Y , an assumption that is upheld when selecting a fixed proportion of the population regardless of what trait is used for selection, $i_X = i_Y$, and the index of relative selection efficiency assumes the form of the ratio of two correlations

$$RSE(X) = \frac{R_{Xy}}{R_{Yy}} . \quad (1)$$

Mass selection

Under mass selection R_{yy} is $\sqrt{h_y}$, R_{xy} is $r\sqrt{h_x}$, where r is the genetic correlation between the two traits and h_x is heritability in the narrow sense of the alternative trait, and the RSE of X is $r\sqrt{h_x/h_y}$. We give this expression the symbol p : i.e.

$$RSE(X) = p = r\sqrt{h_x/h_y} \quad (2)$$

Estimates of p in any particular instance are usually derived from estimates of its component parameters. If the latter are obtained from an analysis of f pairs of parent-progeny records with one progeny per parent, the heritabilities being estimated as twice progeny on parent regressions and the genetic correlation as the ratio of geometric means of appropriate covariances (Hazel, 1943), the estimate of p is

$$\hat{p} = \sqrt{\frac{\hat{\sigma}_{x_o y_p} \hat{\sigma}_{x_p y_o} \hat{\sigma}_{y_p}^2}{\hat{\sigma}_{y_o y_p}^2 \hat{\sigma}_{x_p}^2}} \quad .$$

$\hat{\sigma}_{x_o y_p}$ is the estimated covariance between the phenotypes X in the offspring and Y in the parent -- with a similar notation for the other terms in \hat{p} . Using the methods developed by Reeve (1955) for obtaining the sampling variance of an estimated genetic correlation, it can be shown that the variance of \hat{p} is approximately

$$f[\text{var}(\hat{p})] = 2(1+2p^2+R^2-4Rp)/h_y^2 + h_x(1-2Rp)/2h_y + \frac{1}{2}p(2R+4Rp^2-2R^2p-3p) \quad (3)$$

where R is the phenotypic correlation between the traits.

Decisions in favor of using an alternative trait are likely to be made when estimates of p are close to unity. It is therefore of interest to consider the standard error of \hat{p} as given by (3), when the true value of p is close to 1.00. Table 1 shows six such examples, assuming estimation from 1000 parent-progeny pairs.

Table 1. Standard Errors of \hat{r} and \hat{p}

h_y	h_x	R	r	s.e. (\hat{r})	p	s.e. (\hat{p})
.2	.4	.9	.9	.04	1.27	.15
.2	.4	.9	.7	.07	.99	.10
.2	.4	.7	.7	.08	.99	.18
.3	.4	.9	.9	.03	1.04	.07
.3	.4	.9	.7	.06	.81	.07
.3	.4	.7	.7	.06	.81	.11

For comparison, the standard errors of \hat{r} are also shown, calculated from the expression given by Reeve (1955). Standard errors of estimates obtained from N parent-progeny pairs are those given in the table multiplied by $\sqrt{1000/N}$. It is seen that even for 1000 pairs the standard errors are relatively large, indicating that estimates of p quite different from true value may well arise, at least in the situations specified in Table 1. And this appears to be true generally. Consequently, it seems that a false conclusion might be drawn quite readily about the value of an alternative trait on the basis of an estimated p-value, unless it were obtained from a very large amount of data.

Although estimates of p appear to have relatively large standard errors in many cases, it is instructive to investigate the conditions under which a trait can be preferred as an alternative, by manipulating the inequality $p > 1$ in terms of the population values, r, h_x and h_y . There are undoubtedly situations in which selection on an alternative trait would be preferred even though p was less than unity, depending on the economic and other advantages of using the alternative. In general though, $p > 1$ indicates that the alternative trait is preferable to the basic trait. We then have $r\sqrt{h_x/h_y} > 1$ from which it is seen that because r is less than 1, h_x must exceed h_y . In fact, by re-writing the inequality as $h_x > h_y/r^2$ we see that the minimum heritability that a trait must have in order to be preferable as an alternative to Y is h_y/r^2 . Furthermore, since $h_x > 1$, the minimum value of r for which an alternative can be preferable is $\sqrt{h_y}$. Hence an alternative trait is preferable only if

$$r > \sqrt{h_y} \quad \text{and} \quad h_x > h_y/r^2 .$$

Table 2 shows values of these minimal requirements. While this table in no way solves the problem of large standard errors in estimates of p , it provides a useful set of standards.

Table 2. Minimum requirements for alternative trait to be preferable

Heritability of basic trait h_y	Minimum genetic correlation between alter- native trait and basic trait r	Genetic correlation of alternative trait with basic trait						
		.4	.5	.6	.7	.8	.9	1.0
		Minimum heritability of alternative trait						
.1	.32	.63	.40	.28	.21	.16	.13	.10
.2	.45		.80	.56	.41	.32	.25	.20
.3	.55			.84	.62	.49	.38	.30
.4	.64				.82	.63	.50	.40
.5	.71					.79	.62	.50
.6	.78					.94	.75	.60
.7	.84						.87	.70
.8	.89						.99	.80
.9	.95							.90

Progeny-testing

We will now consider the case of estimating y , the additive genetic merit of an animal (very frequently a sire) from the production records of a group of progeny. For simplification it will be assumed that there is only one record per progeny. As before, the basic objective is improvement in y , in this case through selection of parents on the basis of estimates of y obtained from progeny phenotypes using the selection index procedure. R_{xy} of equation (1) now represents the correlation between an animal's additive genetic merit, y , and an estimate of it based on progeny records of the alternative trait, X ; and R_{yy} is the correlation between y and an estimate thereof based on progeny records of the basic trait Y . When there are N_x progeny records on X

$$R_{xy} = r \sqrt{\frac{N_x h_x}{4 + h_x (N_x - 1)}} \quad (4)$$

while with N_y progeny records on Y

$$R_{yy} = \sqrt{\frac{N_y h_y}{4+h_y(N_y-1)}} \quad (5)$$

The upper limit of R_{yy} for infinite N_y is unity, but that of R_{xy} as N_x tends to infinity is r . Therefore since $r < 1$, there is some value for N_y that gives $R_{yy}=r$, namely

$$N_0 = \frac{r^2(4-h_y)}{h_y(1-r^2)} \quad (6)$$

and for N_y greater than this, R_{xy} is always less than R_{yy} . Hence a progeny-test using the basic trait with N_0 or more progeny is always better than one using the alternative trait, no matter how many progeny are available with records on the latter. [We will use the terms "better" (and "poorer") in comparing the value of traits for selection purposes, in the sense of selection bringing faster (or slower) improvement in y.]

The two traits will be equally efficient when $R_{xy}=R_{yy}$, and this occurs when

$$N_x = \frac{N_y(4-h_y)}{h_x(1-r^2)(N_0-N_y)} \quad (7)$$

Since this expression represents a number of progeny it must be positive and finite. Therefore R_{xy} can equal R_{yy} only when $N_y < N_0$, so that N_0 represents the maximum number of progeny using the basic trait for which it is possible to have a progeny-test using the alternative trait that is equally efficient. And when N_y is less than N_0 , the number of progeny required with records on the alternative trait is N_x , obtainable from (7) in terms of N_y . An indication of these results has been given previously in Searle (1961).

Generally speaking, one would expect N_x as given by (7) to be greater than N_y , and this is usually the case; but it is less than N_y if N_y is less than

$$N_L = N_0 - \frac{4-h_x}{h_x(1-r^2)} \quad (8)$$

By substituting for N_0 from (6) and using $p = r\sqrt{h_x/h_y}$ we find that

$$N_L = 1 + \frac{4(p^2-1)}{h_x(1-r^2)} \quad (9)$$

Again, N_L must be positive, and thus it exists only for $p > 1$. Hence when $p > 1$ and $N_y < N_L$, equivalence with a progeny-test based on N_y progeny using the basic trait can be achieved with actually less than N_y progeny using the alternative trait. Thus we have progeny-tests based on N_x progeny using X and N_y progeny using Y being equivalent for N_x and N_y satisfying equation (7) provided $N_y < N_0$; N_x usually exceeds N_y , but when $p > 1$ and $N_y < N_L$, N_x is less than N_y .

Example: Suppose $h_x=0.6$, $h_y=0.2$ and $r=0.9$. Then from (6) $N_0=81$, so that with more than 81 progeny Y is always preferable to X; but for N_y less than 81, N_x progeny using X are equivalent to N_y progeny using Y where, from (7), $N_x=29.8 N_y/(81-N_y)$. Equation (8) gives $N_L=51.2$, and since $p = 1.56 > 1$, N_x is less than N_y for N_y less than 52; e.g. for $N_y=27$, $N_x=14.9$, which would be taken as 15 since N_x and N_y must be integers.

The above example is illustrated in Figure 1. It is perhaps atypical in that it represents a situation of two traits having quite different heritabilities but a relatively high genetic correlation, it being more customarily found that traits have either dissimilar heritabilities and a low genetic correlation or similar heritabilities and a relatively high genetic correlation. Nevertheless, the figure serves as a useful illustration of the behaviour of R_{V_y} and R_{X_y} as discussed. The curve of R_{X_y} plotted against number of progeny is above that of R_{V_y} for less than N_L progeny, the two curves intersect at the point of N_L progeny, and thereafter the R_{X_y} curve is below the R_{V_y} curve. Between N_L and N_0 progeny, R_{V_y} exceeds R_{X_y} for any given number of progeny, but for a given value of R_{V_y} , R_{X_y} can be found equal to R_{V_y} for a larger number of progeny. For N_0 progeny, R_{V_y} equals the limiting value of R_{X_y} and for this number of progeny and more, the progeny-test using Y is always better than one using X.

As we have seen, the curve of R_{X_y} plotted against number of progeny is above that of R_{V_y} for less than N_L progeny. Hence for a given number of progeny, n say, less than N_L , the progeny-test using the alternative trait will be better

than that using the basic trait on the same number of progeny, i.e. R_{xy}/R_{yy} greater than unity. This implies

$$p > \sqrt{\frac{4+h_x(n-1)}{4+h_y(n-1)}} \quad (10)$$

or, equivalently,

$$r > \sqrt{\frac{4h_y+h_xh_y(n-1)}{4h_x+h_xh_y(n-1)}} \quad (11)$$

Although the existence of N_L and therefore of $n < N_L$ requires $p > 1$, expression (10) provides the exact lower limit on p for any particular n , h_x and h_y .

Similarly, expression (11), although equivalent to (10), provides a lower limit on the genetic correlation for given n , h_x and h_y .

Combinations of traits

The relative selection efficiency of an index I is, by analogy with equation (1),

$$RSE(I) = R_{Iy}/R_{yy} \quad (12)$$

Similarly, the selection efficiency of an index I relative to another index I^* is

$$\frac{RSE(I)}{RSE(I^*)} = \frac{R_{Iy}}{R_{I^*y}} \quad (13)$$

The relative selection efficiency of indices that are combinations of alternative traits will now be considered, comparing each index with Y , and the indices with one another. Thus if the index is

$$I = Y + bX \quad (14)$$

where b is obtained by the usual selection index procedure, it can be shown that

$$RSE(I) = \sqrt{\frac{1+p^2-2Rp}{1-R^2}} \quad (15)$$

where R is the phenotypic correlation between X and Y . As expected, this expression is always greater than unity, even for p less than unity including negative values of p . Hence I is always better than Y , and no matter how poor X is on its own as an alternative to Y , it contributes something to the index relative to using just Y . The important question is how much. It turns out that a very poor X (small values of p) contributes a worthwhile amount only when the phenotypic correlation, R , is large -- and a moderately poor X does so only when R is small. Further, since $RSE(I) > p$, I is better than X even when X is better than Y , $p > 1$, but again the question is to what extent. This is indicated by

$$\frac{RSE(I)}{RSE(X)} = \sqrt{\frac{1+1/p^2-2R/p}{1-R^2}}$$

which is simply $RSE(I)$ with $1/p$ replacing p .

Suppose now that two alternative traits are available, having phenotypes X_1 and X_2 , heritabilities h_1 and h_2 , genetic correlations r_1 and r_2 and phenotypic correlations R_1 and R_2 with the basic trait, and genetic and phenotypic correlations with each other or r_{12} and R_{12} . The simplest comparison between the two alternatives is to find when one, X_1 say, is better than the other, X_2 . This occurs when $RSE(X_1) = p_1 = r_1\sqrt{h_1/h_y}$ exceeds $RSE(X_2) = p_2 = r_2\sqrt{h_2/h_y}$, i.e. when $r_1/r_2 > \sqrt{h_2/h_1}$. Another possibility is that X_1 may be better than I_2 , the index combining X_2 and Y in the manner of equation (14). This will arise when $p_1 > RSE(I_2)$, i.e. when

$$p_1 > \sqrt{\frac{1+p_2^2-2p_2R_2}{1-R_2^2}} .$$

The two alternative traits can also be compared when each is used in an index with Y . Thus I_1 is better than I_2 when $RSE(I_1) > RSE(I_2)$ which reduces to

$$\frac{p_1-R_1}{p_2-R_2} > \sqrt{\frac{1-R_1^2}{1-R_2^2}} .$$

Finally we might consider combining the two alternatives into an index

$$I_x = b_1 X_1 + b_2 X_2$$

and this can be shown to have relative selection efficiency

$$RSE(I_x) = \sqrt{\frac{p_1^2 - 2p_1 p_2 R_{12} + p_2^2}{1 - R_{12}^2}} .$$

It is easily shown that this expression is greater than both p_1 and p_2 and therefore exceeds unity when either or both of these do, thus demonstrating the expected results that I_x is better than either X_1 or X_2 and that it is better than Y when either or both of them are. How much better than X_1 , for example, can be assessed from

$$\frac{RSE(I_x)}{RSE(X_1)} = \sqrt{\frac{1 + (p_2/p_1)^2 - 2R_{12}(p_2/p_1)}{1 - R_{12}^2}} ,$$

which is expression (15) with p_2/p_1 replacing p and R_{12} replacing R .

Four comparisons among two alternative traits have been briefly considered, namely X_1 against X_2 , X_1 against I_2 , I_1 against I_2 and I_x against X . Comparisons with the index $b_1 X_1 + b_2 X_2 + b_3 Y$ could also be made as well as with indices involving three or more alternatives, but increased algebraic complexity makes it difficult to set up useful conditions under which different alternatives are to be preferred. The expressions given above for two alternatives yield a little to further manipulation but their most useful forms are as presented, dependent largely on values of $RSE(I)$ of equation (15). Some progress can also be made in developing expressions for comparing two different alternatives in progeny-test selection.

Literature Cited

- Falconer, D. S., (1960), "Introduction to quantitative genetics," Oliver and Boyd.
- Hazel, L. N., (1943), The genetic basis for constructing selection indexes, Genetics, 28:476-490.
- Lerner, I. M. and Cruden, D. M., (1948), The heritability of accumulative monthly and annual egg production, Poultry Sci., 27:67-77.
- Reeve, E. C., (1955), The variance of the genetic correlation coefficient, Biometrics, 11:357-374.
- Searle, S. R., (1961), Part lactations. III. Progeny-testing with part lactation records, J. Dairy Sci., 44:921-927.
- Young, S. S. Y. and Weiler, H., (1960), Selection for two correlated traits by independent culling levels, J. of Genetics, 57:329-338.

$$\text{Correlations } R_{yy} = \sqrt{\frac{Nh_y}{4+(N-1)h_y}} \quad \text{and} \quad R_{xy} = r \sqrt{\frac{Nh_x}{4+(N-1)h_x}}$$

for $h_y = 0.2$, $h_x = 0.6$, and $r = 0.9$

